Eurographics Conference on Visualization (EuroVis), Posters Track (2016)
T. Isenberg and F. Sadlo (Guest Editors)

Understanding Networks beyond Overlapping Clusters

V. Guchev*, S. Angelini[1 §], G. Amati[1 ‖]

[1]Fondazione Ugo Bordoni, Rome, Italy

Abstract

Tasks associated with the investigation of large complex clustered networks are widespread in various research areas. Among the popular and common approaches to exploratory analysis, it is definitely worthwhile to underscore the node-link-based graph visualization. However, despite the prevalence of node-link-based tools, its graphic design and geometric representation of topology almost invariably formed by a spontaneous spatial structure, or on the contrary, by a too rigidly ordered arrangement. Transformation possibilities of multivariate data structures may allow finding a suitable graphic balance between optic chaos and visual primitiveness by the use of partially ordered sets for grouping. By taking the studying of Twitter communities as a task, the paper presents a data modelling method in conjunction with a set of visualization techniques, which implement a convenient and perceivable interactive toolset for analytical exploration of overlapping network clusters.

1. Introduction

There is a wide range of specialized approaches to the interactive data exploration and analysis of multivariate networks with overlapping clusters, applied in various fields of bioinformatics and geoinformatics, as well as in a number of challenges of the transport, telecommunication, financial and social networks. Quality and efficiency of the expert's investigation activity are reliant on analytical tools, which form a mnemonic abstract representation of the studying phenomenon. Usually the designers of visual interfaces have two base controversial objectives: on the one hand to find ways to use the human potential of heuristic thinking [ZSN*06] in order to generate insights and discoveries [SSKK14], on the other hand to provide systematized symbolic simulation along with ordered topology of graphical representation [SJUS08].

This paper presents the approach and techniques for the visual exploratory analysis of Twitter communities. The input data is a mention graph [AAB*15]: that is a set of profiles of the individuals, who unilaterally or reciprocally mention each other. Moreover, every individual affiliated to one or more "club of interests", that is a community [YL13]. In simple terms, there is a network of users, either of which represented by communicative (via sent and/or received mentions) and affiliative (via one or more joined communities) components.

Objective of the work is developing of a tool that would efficiently assist the analyst to obtain required characteristics of each community or individual, from the studying data extraction. Specifically, the following features should be evidently identified:

- The most active individuals regarding the communicative (i.e. having the largest number of mentions) or affiliative (i.e. joining the largest number of communities simultaneously) aspects.
- Popularity of simultaneous affiliations (all individuals with the identic combination of joined communities form a sub-community).
- Distribution of population among the available communities and sub-communities.
- Patterns of sub-communities expansion: which among sub-communities have most similar affiliation profiles.
- Prevalence of communicative relations among the individuals of similar by affiliation profiles sub-communities.
- Dependencies between the similarity of sub-communities and the character of activity between their members.

In general, the resulting interactive representation should allow to identify patterns and anomalies in affiliative relations and communicative links of the network, as well as to answer the question whether the similar sub-communities form a closely connected narrow circle of individuals. In addition, it should provide the details on request for each type of displaying object: starting with profile attributes of selected individual, ending with the descriptive statistical information about an interesting group. The experimental input dataset is a hypergraph, which comprises around 1 million of vertices (representing profiles of Twitter users) and edges (representing mentions) together with about 100 hyperedges (representing communities) having up to 30 overlappings. By applying filters and selections, the network visualization performed in parts of size from 10 thousand to 100 thousand elements.

* e-mail: guchev.v@gmail.com
§ e-mail: sangelini@fub.it
‖ e-mail: gba@fub.it

© 2016 The Author(s)
Eurographics Proceedings © 2016 The Eurographics Association.

DOI: 10.2312/eurp.20161147

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

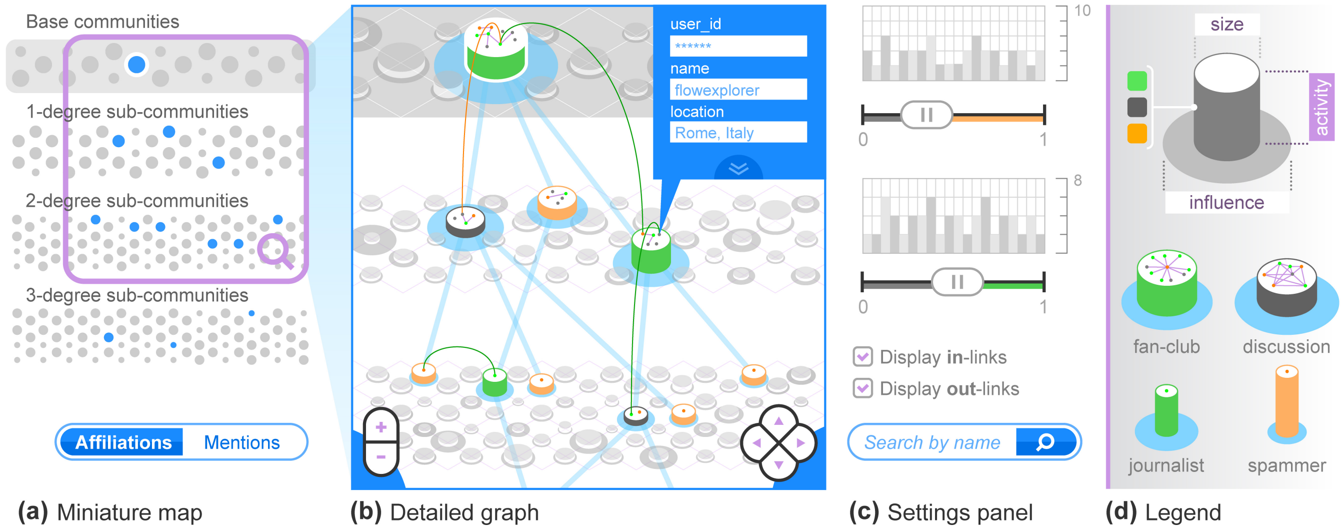

Figure 1: *The interface highlights sub-communities, related to the selected base community. The colors describe the predominance of receiving (green) or sending (orange) mentions. Brim size on each hat glyph is proportional to the number of relatives by affiliation profile.*

2. Related work

Processing of the hypergraph data is typical for a number of studies [Vaz09] [TWS*14], including the following types of research work: knowledge bases for artificial neural networks, biological networks, genome datasets, text analysis. Capabilities of the hypergraph partitioning and clustering by planarization provide wide opportunities for modelling of desired structures with needed hierarchical dependencies [FLR97] [HC15] [BCPS10]. Nevertheless, with few exceptions [SSK14], such data structures are not practically used for the tools of visual analysis. There are some works associated with the graphical use of partially ordered sets: although it suggests new design ideas, it is not going into the issues of application for advanced visualization tools [Pat14] [FAB14]. It is should be noted, that what first comes to mind is Euler diagram, is not applicable for drawing readable graphs even with a moderate hyperedge overlapping degree [SA08]. The most common techniques for improvement and facilitation of the hypergraph exploration (along with browsing of overlapping communities) are the following:

- Visual compression of vertices, for instance the power graph concept [Ahn13] or usage of the motifs [vLKS*11].
- Layered and multilevel spatial representation [BETT99].
- Hybrid approaches for space-filling [IMMS09].

The most of existing tools and techniques for drawing of node-link diagrams, which in theory should allow understanding networks beyond overlapping clusters, have several drawbacks: generally, it is the lack of habitual for the human geometric order in the composition of elements, together with chaotic and irreproducible pattern for every single (even resembling) dataset. Overall, taking into account the specifics of social networks (which may comprise thousands of related overlapped clusters), it became obvious that this particular case of research requires a unique complex solution.

3. Design

Since the core of work is the exploration of a large amount of groups (which formed by intersection of overlapping communities) together with individuals-members of these groups, the primary function of the interface is the arranging and demonstration of these groups in accordance with the available inter-communal relations and the existing mention-links between individuals. For ease of reading the design includes the screen grid, handy for context layout of additional statistical charts. The data representation way used in the design influenced by the idea of partially ordered sets: which may be constructed in different variations of lattice categories, such as lattice or semi-lattice, or sub-lattice, or congruence, or morphisms, etc. One of the major tasks of the interface is the finding of sub-communities related to certain base communities, or vice-versa, the detection of base communities forming a sub-community. The lattice concept allows distribution of the groups through layers and provides the both of the intralayer ordering and flexible forming of inter-communal relations on the base of required criterion.

As a source of inspiration, the interface architecture has taken concepts of the coordinated views and the context highlighting of patterns, in combination with principles of the overview and detail and the parallel coordinates. The given interface in Figure 1 composed of three main blocks: the miniature map, the detailed graph view of a map section, the panel of context tools and settings. A feature of the design is a pseudo-spatial glyph in the shape of cylinder hat. Modulation of the hat size and color allows displaying multiple values, the simple geometry minimizes optical illusions at the collocation. The interface is initially oriented on a productive computing power and a large high-resolution display.

Acknowledgements. This work has been carried out in the MiSE-ISCOM laboratories within a long-term research agreement between FUB and ISCOM.

© 2016 The Author(s)
Eurographics Proceedings © 2016 The Eurographics Association.

References

[AAB*15] AMATI G., ANGELINI S., BIANCHI M., FUSCO G., GAMBOSI G., GAUDINO G., MARCONE G., ROSSI G., VOCCA P.: Moving beyond the twitter follow graph. In *KDIR 2015 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015), Volume 1, Lisbon, Portugal, November 12-14, 2015* (2015), Fred A. L. N., Dietz J. L. G., Aveiro D., Liu K., Filipe J., (Eds.), SciTePress, pp. 612–619. URL: http://dx.doi.org/10.5220/0005616906120619, doi:10.5220/0005616906120619. 1

[Ahn13] AHNERT S. E.: Power graph compression reveals dominant relationships in genetic transcription networks. *Mol. BioSyst. 9* (2013), 2681–2685. URL: http://dx.doi.org/10.1039/C3MB70236G, doi:10.1039/C3MB70236G. 2

[BCPS10] BRANDES U., CORNELSEN S., PAMPEL B., SALLABERRY A.: Blocks of hypergraphs - applied to hypergraphs and outerplanarity. In *IWOCA* (2010), Iliopoulos C. S., Smyth W. F., (Eds.), vol. 6460 of *Lecture Notes in Computer Science*, Springer, pp. 201–211. URL: http://dblp.uni-trier.de/db/conf/iwoca/iwoca2010.html#BrandesCPS10a. 2

[BETT99] BATTISTA G. D., EADES P., TAMASSIA R., TOLLIS I. G.: *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice-Hall, 1999. 2

[FAB14] FATTORE M., ARCAGNI A., BARBERIS S.: Visualizing Partially Ordered Sets for Socioeconomic Analysis. *Revista Colombiana de Estadistica 37*, 2 (2014), 437–450. 2

[FLR97] FEINBERG V., LEVIN A., RABINOVICH E.: *VLSI Planarization: Methods, Models, Implementation*. Mathematics and its applications. Kluwer Academic, 1997. URL: https://books.google.it/books?id=MzFTAAAAMAAJ. 2

[HC15] HEINTZ B., CHANDRA A.: Enabling scalable social group analytics via hypergraph analysis systems. In *7th USENIX Workshop on Hot Topics in Cloud Computing, HotCloud '15, Santa Clara, CA, USA, July 6-7, 2015.* (2015). URL: https://www.usenix.org/conference/hotcloud15/workshop-program/presentation/heintz. 2

[IMMS09] ITOH T., MUELDER C., MA K.-L., SESE J.: A hybrid space-filling and force-directed layout method for visualizing multiple-category graphs. *2014 IEEE Pacific Visualization Symposium 0* (2009), 121–128. doi:http://doi.ieeecomputersociety.org/10.1109/PACIFICVIS.2009.4906846. 2

[Pat14] PATTISON T.: Interactive visualization of formal concept lattices. In *Joint Proceedings of the Fourth International Workshop on Euler Diagrams and the First International Workshop on Graph Visualization in Practice co-located with Diagrams 2014, Melbourne, Australia, July 28th and 1st August 2014.* (2014), Burton J., Stapleton G., Klein K., (Eds.), vol. 1244 of *CEUR Workshop Proceedings*, CEUR-WS.org, pp. 78–89. URL: http://ceur-ws.org/Vol-1244/GViP-paper6.pdf. 2

[SA08] SIMONETTO P., AUBER D.: Visualise Undrawable Euler Diagrams. In *12th International Conference on Information Visualisation* (London, United Kingdom, July 2008), pp. 594–599. URL: https://hal.archives-ouvertes.fr/hal-00319119, doi:10.1109/IV.2008.78. 2

[SJUS08] SCHULZ H.-J., JOHN M., UNGER A., SCHUMANN H.: Visual Analysis of Bipartite Biological Networks. In *Eurographics Workshop on Visual Computing for Biomedicine* (2008), Botha C., Kindlmann G., Niessen W., Preim B., (Eds.), The Eurographics Association. doi:10.2312/VCBM/VCBM08/135-142. 1

[SSK14] SAKET B., SIMONETTO P., KOBOUROV S.: Group-Level Graph Visualization Taxonomy. In *EuroVis - Short Papers* (2014), Elmqvist N., Hlawitschka M., Kennedy J., (Eds.), The Eurographics Association. doi:10.2312/eurovisshort.20141162. 2

[SSKK14] SACHA D., SENARATNE H., KWON B. C., KEIM D. A.: Uncertainty Propagation and Trust Building in Visual Analytics. *Accepted at IEEE VIS 2014 - Provenance for Sensemaking Workshop (poster paper)* (2014). 1

[TWS*14] TAO H., WU Z., SHI J., CAO J., YU X.: Overlapping community extraction: A link hypergraph partitioning based method. In *IEEE International Conference on Services Computing, SCC 2014, Anchorage, AK, USA, June 27 - July 2, 2014* (2014), pp. 123–130. URL: http://dx.doi.org/10.1109/SCC.2014.25, doi:10.1109/SCC.2014.25. 2

[Vaz09] VAZQUEZ A.: Finding hypergraph communities: a bayesian approach and variational solution. *Journal of Statistical Mechanics: Theory and Experiment 2009*, 07 (2009), P07006. URL: http://stacks.iop.org/1742-5468/2009/i=07/a=P07006. 2

[vLKS*11] VON LANDESBERGER T., KUIJPER A., SCHRECK T., KOHLHAMMER J., VAN WIJK J. J., FEKETE J., FELLNER D. W.: Visual analysis of large graphs: State-of-the-art and future research challenges. *Comput. Graph. Forum 30*, 6 (2011), 1719–1749. URL: http://dx.doi.org/10.1111/j.1467-8659.2011.01898.x, doi:10.1111/j.1467-8659.2011.01898.x. 2

[YL13] YANG J., LESKOVEC J.: Overlapping community detection at scale: A nonnegative matrix factorization approach. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* (New York, NY, USA, 2013), WSDM '13, ACM, pp. 587–596. URL: http://doi.acm.org/10.1145/2433396.2433471, doi:10.1145/2433396.2433471. 1

[ZSN*06] ZUK T., SCHLESIER L., NEUMANN P., HANCOCK M. S., CARPENDALE S.: Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (New York, NY, USA, 2006), BELIV '06, ACM, pp. 1–6. URL: http://doi.acm.org/10.1145/1168149.1168162, doi:10.1145/1168149.1168162. 1

© 2016 The Author(s)
Eurographics Proceedings © 2016 The Eurographics Association.